ECONOMETRICS I, Assignment I

Ruoheng Du

02/23/2023

Question 1

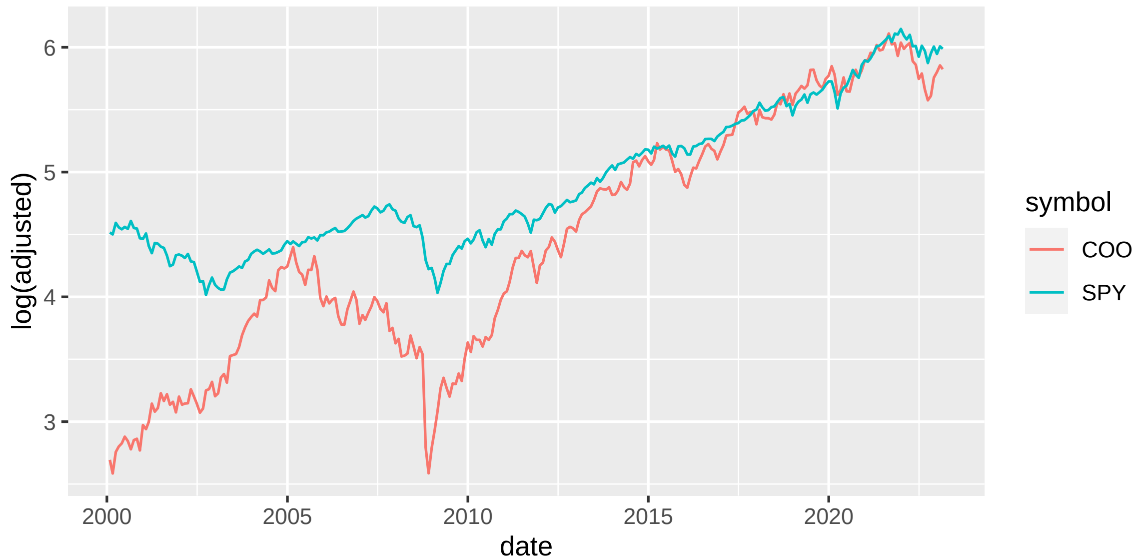
```
library(tidyquant)
library(dplyr)

mydata <- tq_get(c("SPY", "COO"), get = "stock.prices", from = "2000-01-01",)
mydata <- mydata %>%
  group_by(symbol) %>%
    tq_transmute(select = adjusted, mutate_fun = to.monthly, indexAt = "lastof") %>%
    tq_mutate(adjusted, periodReturn, col_rename = "return")

library(ggplot2)

## HERE CODE TO MAKE THE PLOTS
##
head(mydata)
```

```
# A tibble: 6 x 4
# Groups:   symbol [1]
  symbol date       adjusted  return
  <chr>  <date>        <dbl>   <dbl>
1 SPY    2000-01-31     91.5  0
2 SPY    2000-02-29     90.1 -0.0152
3 SPY    2000-03-31     98.8  0.0969
4 SPY    2000-04-30     95.3 -0.0351
5 SPY    2000-05-31     93.8 -0.0157
6 SPY    2000-06-30     95.7  0.0197
```

```
tail(mydata)
```

```
# A tibble: 6 x 4
# Groups:   symbol [1]
  symbol date       adjusted  return
  <chr>  <date>        <dbl>   <dbl>
1 COO    2022-09-30     264. -0.0819
2 COO    2022-10-31     273.  0.0360
3 COO    2022-11-30     316.  0.157
4 COO    2022-12-31     331.  0.0453
5 COO    2023-01-31     349.  0.0553
6 COO    2023-02-28     338. -0.0300
```

```
ggplot(mydata, aes(date, log(adjusted), color = symbol)) +
  geom_line()
```




```
ggplot(mydata, aes(date, return, color = symbol)) +
  geom_line()
```

DISCUSSION
1. For the plot of the date and the logarithm of the price, we can expect that SPY and COO do have some positive correlation, but the extent to which is not that big as we can see that before the financial crisis, the logarithm of the prices of SPY and COO move reversely: while one is declining, the other is increasing. And between 2010 and 2015, they do show the same pattern, but their rates of increase differ a lot. Besides, the logarithm of the prices of SPY and COO have a smaller difference after 2015. 2. For the plot of the date and the return, we can find that both COO and SPY have a lot of spread of returns as they constantly fluctuate around 0. But compared with the SPY, COO has larger and more frequent extreme negatives, which indicates that it may have a larger left skewness.

Question 2

```
library(fBasics)
library(knitr)
```

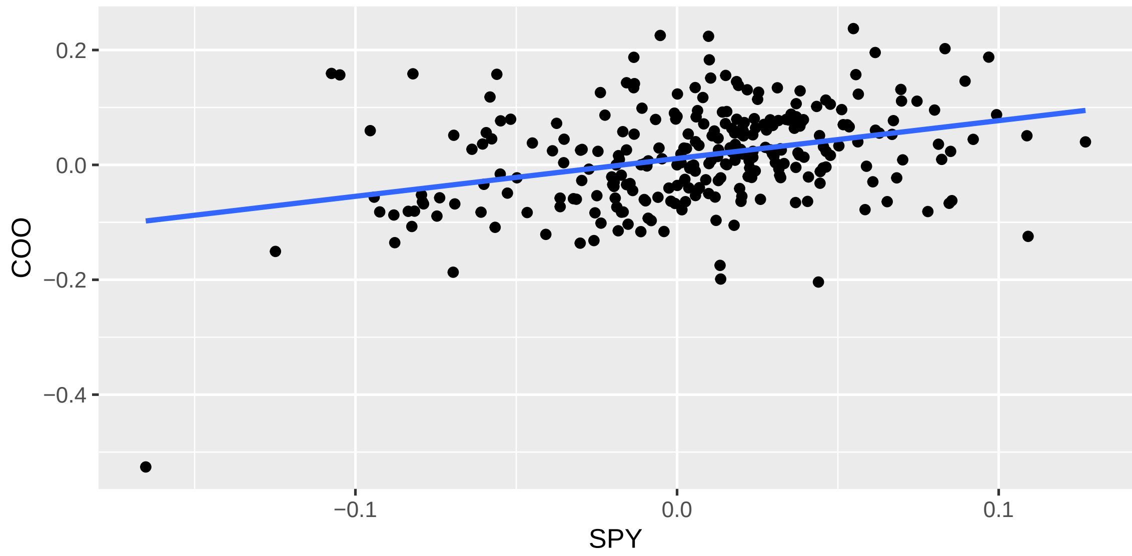
```
mydata %>%
  group_by(symbol) %>%
  summarize(Average = mean(return), StdDev = sd(return),
            Skewness = skewness(return), Kurtosis = kurtosis(return)) %>% # INSIDE SUMMARIZE YOU NEED T
  kable(digits = 3)                          # THE STD DEV, SKEWNESS, KURTOSIS
```

symbol	Average	StdDev	Skewness	Kurtosis
COO	0.015	0.086	-0.784	4.700
SPY	0.006	0.045	-0.465	0.715

DISCUSSION 1. For COO, the average return is 0.015 with a standard deviation of 0.086. This means that a range from -0.157 to 0.187 includes around 95% of the sample, if the observations follow a normal distribution. The skewness of -0.784 implies that the return of COO have a longer tail than others with skewness of 0. And the negative sign indicates that the distribution of the return of COO is left-skewed and has more extreme negative. The excess kurtosis of the return of COO is 4.700, which means that the return of COO has a heavy tail and contains more possibilities to extremes. 2. For SPY, the average return is 0.006 with a standard deviation of 0.045. Compared with COO, SPY has a lower average return and a lower squared spread of the distribution. The skewness of -0.465, which is lower than that of COO, indicates that SPY has a left tail and it is less left-skewed than COO. The excess kurtosis of the return of SPY is 0.715. This is close to normal distribution, which has 0 excess kurtosis, and it is much less than that of COO, indicating that SPY has less possibility for extremes.

Question 3

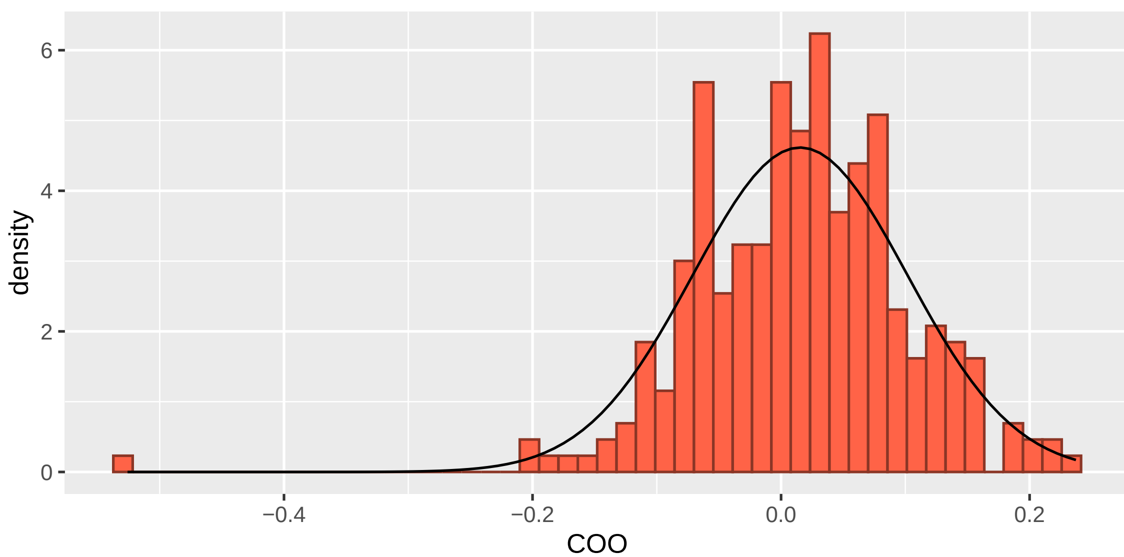
```
mydata.wide <-
  mydata %>%
  select(date, symbol, return) %>%
  tidyr::spread(symbol, return)

## ADD HERE PLOTTING
##
ggplot(mydata.wide, aes(SPY, COO)) +
  geom_point() +
  geom_smooth(method="lm", se=FALSE)
```



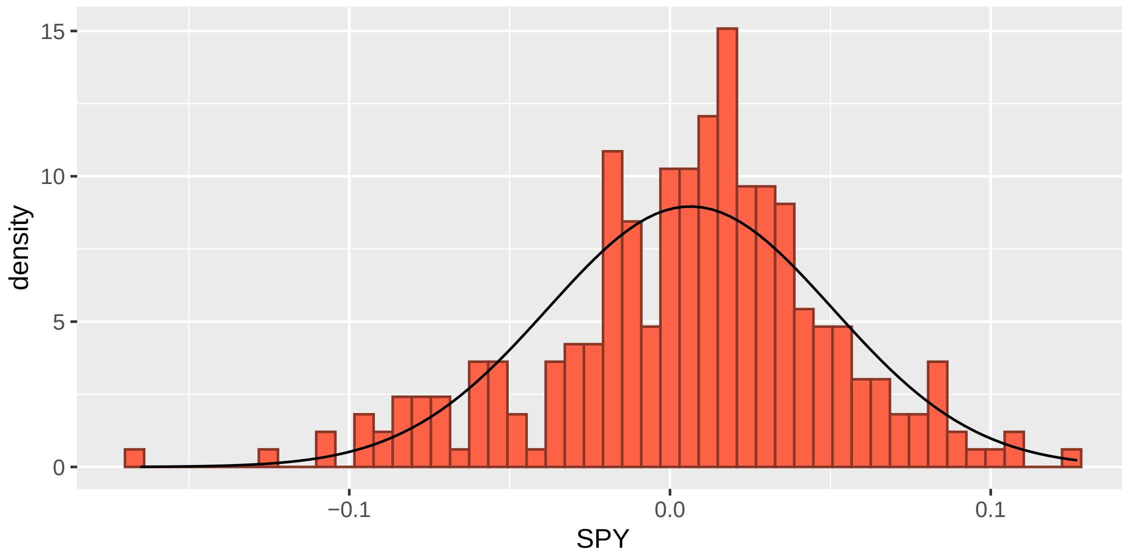
DISCUSSION We can conclude from the scatter plot and the regression line that there is some degree of dependence between the SPY and COO, as the data has a slight upward sloping distribution pattern and the slope of the regression line is indeed positive. Yet, it is uncertain if the two variables are strongly related or whether evidence of a strong effect exists because: 1. there's a number of caveats (for example, the point in the lower left corner) and 2. there are few points on the regression line, but there are many points scattered around.

Question 4

```
##
## PLOT HISTOGRAM ...
##
ggplot(mydata.wide, aes(COO)) +
  geom_histogram(aes(y = ..density..),
                 bins = 50, fill = "tomato1", color = "tomato4") +
  stat_function(fun = dnorm,
                args = list(mean = mean(mydata.wide$COO,na.rm=T),
                            sd = sd(mydata.wide$COO,na.rm=T)))
```




```
ggplot(mydata.wide, aes(SPY)) +
  geom_histogram(aes(y = ..density..),
                 bins = 50, fill = "tomato1", color = "tomato4") +
  stat_function(fun = dnorm,
                args = list(mean = mean(mydata.wide$SPY,na.rm=T),
                            sd = sd(mydata.wide$SPY,na.rm=T)))
```



DISCUSSION
1. For COO, we can see that it has a heavy and long left tail compared with normal distribution. And as we can seen from the histogram, some of the parts lie above the normal distribution line (more frequency), while others lie below the normal distribution line (less frequency). This indicates that for some certain ranges of return, COO has more probability to give these returns than the normal distribution predicts, while for some other ranges of return, this histogram has less probability to give these returns (for instance, between -0.2 and -0.1). 2. For SPY, we can see that it still has a left tail, but this is not as heavy as COO since the left extreme of SPY is not as extreme as that of COO. There's also some parts in the histogram that have less or more frequency than the normal distribution. For instance, between around -0.05 and -0.02, this histogram has less probability to give these returns than the normal distribution predicts.

Question 5

```
mydata.wide %>% select(SPY,COO) %>% cor(use = "complete") %>% kable(digits = 3)
```

	SPY	COO
SPY	1.000	0.339
COO	0.339	1.000

```
## ADD CALCULATION OF THE OLS ESTIMATE OF BETA1
s_xy      = cov(mydata.wide$SPY, mydata.wide$COO)
s2_x      = var(mydata.wide$SPY)
beta1_hat = s_xy / s2_x
beta1_hat
```

```
[1] 0.6588185
```

DISCUSSION The correlation between SPY and COO is 0.339, which indicate a relatively moderate positive linear relationship between SPY and COO. Intuitively, this means that SPY and COO tend to increase at the same time.

Question 6

```
##
## Regression of stock returns on market returns
```

```r
##
fit <- lm(COO ~ SPY, data = mydata.wide)
summary(fit)
```

```
Call:
lm(formula = COO ~ SPY, data = mydata.wide)

Residuals:
     Min       1Q   Median       3Q      Max
-0.42822 -0.04834 -0.00508  0.04826  0.21884

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.011152   0.004934   2.260   0.0246 *
SPY         0.658819   0.109932   5.993 6.42e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.08145 on 276 degrees of freedom
Multiple R-squared:  0.1151,    Adjusted R-squared:  0.1119
F-statistic: 35.92 on 1 and 276 DF,  p-value: 6.419e-09
```

DISCUSSION From lm(), we can conclude that the intercept (beta_0) is equal to 0.011152 and the SPY coefficient (beta_1) is 0.658818. 1. For the intercept/beta_0, it means that if the return of SPY equals zero, the predicted/estimated expected return of COO is 0.011152. Since it is practical to have the return of SPY to be zero, the intercept is economically meaningful. 2. For the slope/beta_1, it means that a one percentage point change in the return of SPY is associated with an estimated expected 0.658818 percentage point change in the return of COO. 3. For the overall fit of the regression, we can learn from the regression that SER = 0.08145 and r_squared = 0.1151. This means that the return of SPY explains for 11.51% for the observed return of COO and that for a given prediction of the return of COO using the return SPY as a predicting variable will on average yield a result which is 0.08145 from the actual return. Although the absolute value of SER is small, this is a relatively large value compared with the standard deviation of the return of COO and SPY, which will make the regression model less predictive.